JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 21, 1-6 (1968)

A Valid Mathematical Model for Approximate Nonlinear Minimal-Variance Filtering*

LAWRENCE SCHWARTZ

Hughes Aircraft Company, Culver City, California 90230

AND

EDWIN B. STEAR

University of California, Los Angeles, California 90024

Submitted by J. M. Richardson

INTRODUCTION

The problem considered herein is the continuous estimation of the state of a physical system undergoing random motion, where the knowledge of the motion is given by a known function of the state plus noise. The physical problem is abstracted to the extent that the random motion is considered to be a Markov process, and the noise is assumed to be white noise; consequently, the resulting mathematical problem is formulated in terms of stochastic differential equations. The optimal estimate for a minimal-variance criterion is known to be the conditional expectation of the state of the system given the measurements; the filter derived herein is a stochastic differential equation for an approximation to the optimal estimate.

A formal derivation of a stochastic differential equation for an approximate nonlinear filter was presented by Bass *et al.* [1], based on an approach suggested by Kushner [2] and developed partially by Bucy [3]. It was later shown by Schwartz and Bass [4] that inherent in the approach is the assumption that the conditional probability density function is nearly completely contained in a neighborhood of the conditional expectation, and a new approximation was derived that replaced that assumption with the assumption that the conditional density is nearly Gaussian. The same filter was independently derived by Fisher [5].

* Partial support was available from the United States Air Force under Contract Numbers AF-04(694) 826 and AFOSR 699-65.

By the validity of the stochastic differential equations is meant their satisfaction of existence and uniqueness conditions, which are given as

THEOREM 0. *Consider the stochastic differential equation*

$$dx = f(t, x)\, dt + g(t, x)\, dw(t), \tag{1}$$

where x is an n-vector, t is a scalar, w is an m-vector of independent Wiener processes for $m \leqslant n$, f is an n-vector, and g is an $n \times m$ matrix. Suppose that $x(t_0)$ is independent of the elements of w and that f and g are defined for $t \in [t_0, t_f]$ and $x \in R^n$ (Euclidian n-space), are measurable with respect to the set of variables, and that they satisfy the following conditions:

1. *For every $C > 0$, there exists an L_C such that*

$$(t_f - t_0)|f(t, x) - f(t, y)|^2 + \sum_{i=1}^{m} |g^i(t, x) - g^i(t, y)|^2 \leqslant L_C{}^2 |x - y|^2$$

if $|x| \leqslant C$ and $|y| \leqslant C$, where $|\cdot|$ denotes the Euclidian norm of the vector and g^i is the ith column of g.

2. *There exists a K at which*

$$|f(t, x)|^2 + \sum_{i=1}^{m} |g^i(t, x)|^2 \leqslant K(|x|^2 + 1).$$

In such a case, (1) *has a continuous solution with probability one; also, if there are two solutions, with probability one both coincide at all points $t \in [t_0, t_f]$.*
The proof follows from Theorem 3, p. 21, and Theorem 4, p. 56, of Skorokhod [6].

It is shown in the present analysis that further modification to the approximation is necessary to insure that the resulting filter equation is rigorously valid.

PROBLEM STATEMENT

The mathematical formulation of the characteristics of the dynamics and the measurement is given in terms of two random processes which are jointly Markovian. The first is x, given by (1), which represents the motion of the system. The second is z, which is related to the measurement as follows: If it is assumed that the measurement made of x is in the form of an ℓ-vector, for $\ell \leqslant n$, denoted by $h(t, x)$ plus an ℓ-vector noise which is a white noise process, an integrated measurement process can be defined by

$$dz = h(t, x)\, dt + r(t)\, db, \tag{2}$$

where r is a nonsingular symmetric $\ell \times \ell$ matrix, and b is an ℓ-vector of independent Wiener processes, also independent of w. Note that r is explicitly a function only of t. The joint process (x, z) is then given by

$$\begin{bmatrix} dx \\ dz \end{bmatrix} = \begin{bmatrix} f(t, x) \\ h(t, x) \end{bmatrix} dt + \begin{bmatrix} g(t, x) & 0 \\ 0 & r(t) \end{bmatrix} \begin{bmatrix} dw \\ db \end{bmatrix}. \tag{3}$$

It is assumed that (3) satisfies the conditions of Theorem 0.

Let $\hat{x}(t)$ denote the optimal estimate of x evaluated at t: i.e., the conditional expectation of $x(t)$ given $z(s)$, $0 \leqslant s \leqslant t$. The fundamental problem is to derive a stochastic differential equation for $\hat{x}$ with dz as a differential forcing function. The exact equation is known to be (See Bass *et al.* [1])

$$d\hat{x} = \hat{f}\, dt + (\widehat{xh^*} - \hat{x}\hat{h}^*)\, r^{-2}(dz - \hat{h}\, dt), \tag{4}$$

where the asterisk is used to denote the transpose. It is not practical to mechanize (4). The previous derivations for an approximation to (4) involved the expansion of the nonlinear functions in a truncated Taylor's series, and neither the approximate dynamical equations nor the resulting filter equations satisfy Theorem 0. The problem considered herein is to derive an approximation to (4) in such a way that the existence and uniqueness conditions are not violated by either set of equations.

Modified Equations

For simplicity, the dependence of the system dynamics on t will not be noted explicitly in this section, since the time variation has no effect on the form of the filter equations. In place of the truncated Taylor's series used previously, the following approximation is introduced, assuming all the derivatives exist, and using the summation convention:

$$f_i^k(x) = f_i(\hat{x}) + e^{-k|x-\hat{x}|^2}[f_{ij}^{(1)}(\hat{x})(x_j - \hat{x}_j) + \tfrac{1}{2} f_{ij\ell}^{(2)}(\hat{x})(x_j - \hat{x}_j)(x_\ell - \hat{x}_\ell)], \tag{5}$$

where f_i is the ith row of f, $f_{ij}^{(1)}$ denotes $\partial f_i / \partial x_j$, $f_{ij\ell}^{(2)}$ denotes $\partial^2 f_i / \partial x_j \partial x_k$ and $|\ |$ denotes the Euclidean norm. For now, k is an arbitrary positive real number. Similarly h^k means the expansion of h in the modified series, while g^k is the expansion of g in the form

$$g_{ij}^k(x) = g_{ij}(\hat{x}) + e^{-k|x-\hat{x}|^2} g_{ij\ell}^{(1)}(\hat{x})(x_\ell - \hat{x}_\ell). \tag{6}$$

The modified joint system is

$$\begin{bmatrix} dx \\ dz \end{bmatrix} = \begin{bmatrix} f^k(x) \\ h^k(x) \end{bmatrix} dt + \begin{bmatrix} g^k(x) & 0 \\ 0 & r(t) \end{bmatrix} \begin{bmatrix} dw \\ db \end{bmatrix}, \tag{7}$$

which clearly satisfies the conditions of Theorem 0.

If $k = 0$, the expansions for f and h are exactly the truncated Taylor's series used previously. The approximation for g, even with $k = 0$, is different: Previously, gg^* was expanded in a second-degree Taylor's series, which may result in an approximation to g which is not real. Since the earlier analyses were not concerned with the approximate system dynamics, *per se*, the nonexistence of g as a real vector was unimportant; this is not the case for the present analysis.

The derivation of the modified filter equations is quite similar to the derivations in Bass *et al.* [1] and Schwartz and Bass [4], and is discussed in some detail in Schwartz [7]. The utility of the modified approximation of (5) and (6) follows from the assumption that the conditional covariance is approximately Gaussian, since the modified equations involve what might be called a weighted central moment tensor with components

$$[\mu_k{}^m(P)]_{i_1 \ldots i_m} \triangleq \int_{R^n} e^{-k|x-\hat{x}|^2} \prod_{j=1}^{m} (x_{i_j} - \hat{x}_{i_j}) \frac{\exp[-\frac{1}{2}(x-\hat{x})^* P^{-1}(x-\hat{x})]}{(2\pi)^{n/2}(\det P)^{1/2}} \, dx. \tag{8}$$

From (8) it follows that

$$\mu_k{}^m(P) = [\det(2kP + I_n)]^{-1/2} \mu_0{}^m[(2kI_n + P^{-1})^{-1}]. \tag{9}$$

Let $P^k \triangleq \mu_k{}^2(P)$, $c_k \triangleq [\det(2kP + I_n)]^{-1/2}$, and

$$T^k_{ijpq} \triangleq (P^k_{ij} - c_k P_{ij}) P^k_{pq} + P^k_{iq} P^k_{pj} + P^k_{ip} P^k_{qj} .$$

It is shown in Schwartz [7] that the modified approximate filter equations are

$$\begin{aligned} d\hat{x}_i \approx\; & f_i(\hat{x})\, dt + \tfrac{1}{2} f^{(2)}_{ij\ell}(\hat{x})\, p^k_{j\ell}\, dt \\ & + P^k_{ij} h^{(1)}_{\ell j}(\hat{x})\, r^{-2}_{\ell m}[dz_m - (h_m(\hat{x}) + \tfrac{1}{2} h^{(2)}_{mnp}(\hat{x})\, P^k_{np})]\, dt. \end{aligned} \tag{10}$$

$$\begin{aligned} dP_{ij} \approx\; & P^k_{i\ell} f^{(1)}_{j\ell}(\hat{x})\, dt + f^{(1)}_{i\ell}(\hat{x})\, P^k_{\ell j}\, dt - P^k_{i\ell} h^{(1)}_{m\ell}(\hat{x})\, r^{-2}_{mn} h^{(1)}_{np}(\hat{x})\, P^k_{pj}\, dt \\ & + g_{i\ell}(\hat{x})\, g_{j\ell}(\hat{x})\, dt + g^{(1)}_{i\ell m}(\hat{x})\, g^{(1)}_{i\ell n}(\hat{x})\, P^k_{mn}\, dt \\ & + \tfrac{1}{2} c_k^{-1} T^k_{ij\ell m} h^{(2)}_{n\ell m}(\hat{x})\, r^{-2}_{np}[dz_p - (h_p(\hat{x}) + \tfrac{1}{2} h^{(2)}_{pqr}(\hat{x})\, P^k_{qr})\, dt], \end{aligned} \tag{11}$$

where P^k is computed from P.

To see that the existence and uniqueness conditions of Theorem 0 can be satisfied for some f, g, and h it is convenient to recast (10) and (11) in the form of a single vector stochastic differential equation for the vector s of dimension $\frac{1}{2}n(n+3)$ with components defined by $s_i = x_i$ for $i = 1,\ldots, n$ and $s_i = P_{pq}$ (where $i = q + np - \frac{1}{2}p(p-1)$) for $q = 1,\ldots, p$; $p = 1,\ldots, n$. The vector equation takes the form

$$ds = a^k(t, s)\, dt + b^k(t, s)\, dz, \tag{12}$$

where the argument t is again explicitly noted, and both a^k and b^k are linear combinations of terms from (10) and (11). In forming a^k and b^k, use is made of the fact that $P_{pq} = P_{qp}$. Explicitly, for $i = n$

$$\begin{aligned} a_i^k(t, s) &= f_i(t, s) + \tfrac{1}{2} f^{(2)}_{ij\ell}(t, s)\, P^k_{j\ell}(s) - P^k_{ij}(s)\, h^{(1)}_{\ell j}(t, s)\, r^{-2}_{\ell m}(t)\, h_m(t, s) \\ &\quad - \tfrac{1}{2} P^k_{ij}(s)\, h^{(1)}_{\ell j}(t, s)\, r^{-2}_{\ell m}(t)\, h^{(2)}_{mnp}(t, s)\, P^k_{np}(s), \end{aligned} \tag{13}$$

$$b_i^k(t, s) = P^k_{ij}(s)\, h^{(1)}_{\ell j}(t, s)\, r^{-2}_{\ell m}(t) \tag{14}$$

and for the remaining elements

$$\begin{aligned} a^k_{ij}(t, s) &= P^k_{i\ell}(s) f^{(1)}_{j\ell}(t, s) + P^k_{\ell j}(s) f^{(1)}_{i\ell}(t, s) \\ &\quad - P^k_{i\ell}(s)\, P^k_{pj}(s)\, h^{(1)}_{\ell m}(t, s)\, h^{(1)}_{np}(t, s)\, r^{-2}_{mn}(t) \\ &\quad + g_{i\ell}(t, s)\, g_{j\ell}(t, s) + P^{2k}_{\ell m}(s)\, g^{(1)}_{in\ell}(t, s)\, g^{(1)}_{jnm}(t, s) \\ &\quad - \tfrac{1}{2} c_k^{-1}(s)\, T^k_{ij\ell m}(s)\, h^{(2)}_{n\ell m}(t, s)\, r^{-2}_{np}(t)\, h_p(t, s) \\ &\quad - \tfrac{1}{4} c_k^{-1}(s)\, T^k_{ij\ell m}(s)\, h^{(2)}_{n\ell m}(t, s)\, r^{-2}_{np}(t)\, h^{(2)}_{pqr}(t, s)\, P^k_{qr}(s), \end{aligned} \tag{15}$$

$$b^k_{ij}(t, s) = \tfrac{1}{2} c_k^{-1}(s)\, T^k_{ij\ell m}(s)\, h^{(2)}_{n\ell m}(t, s)\, r^{-2}_{np}(t). \tag{16}$$

It is easily verified that $c_k^{-1} T^k$ satisfies condition 2 of Theorem 0 and that P^k is bounded for non-negative definite P. By inspection, from (13)–(16), it then follows that the conditions of Theorem 0 applied to (12) are satisfied if the following hold:

1. P is non-negative definite

2. f is twice continuously differentiable such that condition 2 of Theorem 0 is satisfied by each f_i, $f^{(1)}_{ij}$, and $f^{(2)}_{ijk}$.

3. h is twice continuously differentiable such that $h^{(1)}$, $h^{(2)}$ and $|h| \cdot |h^{(2)}|$ are bounded.

4. g is continuously differentiable such that condition 2 of Theorem 0 is satisfied by each $g_{ij} g_{k\ell}$ and $g^{(1)}_{ijk}\, g^{(1)}_{\ell mn}$.

References

1. R. W. Bass, V. D. Norum, and L. Schwartz. Optimal multichannel nonlinear filtering. *J. Math. Anal. Appl.* **16** (1966), 152-164.
2. H. J. Kushner. On the differential equations satisfied by conditional probability densities of Markov processes. *J. SIAM Contr. Ser. A.*, **2** (1964), 106-119.
3. R. S. Bucy. Nonlinear filtering theory. *IEEE Trans. Automatic Contr.* AC-10 (1965), 198.
4. L. Schwartz and R. W. Bass. Extensions to optimal multichannel nonlinear filtering. Hughes Aircraft Co., Culver City, California, Report SSD 60220 R, Feb. 21, 1966.
5. J. R. Fisher. Conditional probability density functions and optimal nonlinear estimation. Ph. D. thesis. University of California, Los Angeles, 1966.
6. A. V. Skorokhod. "Studies in the Theory of Random Processes." Reading, Mass., Addison-Wesley, 1965.
7. L. Schwartz. Approximate continuous nonlinear minimal-variance filtering. Ph. D. thesis. University of California, Los Angeles, 1966.